

1 論文メモ書き

1.1 Momentum SGD

非同期な SGD の実装は”A lock-free approach to parallelizing stochastic gradient descent”などで検討されていて、実装が簡単.

θ を全てのスレッド間で共有するパラメータベクトル, $\Delta\theta_i$ を i 番目のスレッドによって計算された θ の勾配とする.

各スレッド i はモメンタム項の更新 $m_i = \alpha m_i + (1 - \alpha)\Delta\theta_i$ とパラメータの更新

$$\theta \leftarrow \theta - \eta m_i$$

をロックなしで独立して行う.

ここで、各スレッドは独自の勾配とモメンタムベクトルを保持している.

1.2 RMSProp

非同期な RMSProp に関する研究はあまりされていない.

標準的な RMSProp の更新式は、以下で与えられる.

$$g = \alpha g + (1 - \alpha)\Delta\theta^2$$

$$\theta \leftarrow \theta - \eta \frac{\Delta\theta}{\sqrt{g + \epsilon}}$$

非同期で RMSProp を適用するには、要素ごとの g の移動平均をスレッドごとに共有するかどうかを決定する必要がある.